# A stochastic density matrix approach to approximation of probability distributions and its application to nonlinear systems[*]

Igor G. Vladimirov[†]

## Abstract

This paper outlines an approach to the approximation of probability density functions by quadratic forms of weighted orthonormal basis functions with positive semi-definite Hermitian matrices of unit trace. Such matrices are called stochastic density matrices in order to reflect an analogy with the quantum mechanical density matrices. The SDM approximation of a PDF satisfies the normalization condition and is nonnegative everywhere in contrast to the truncated Gram-Charlier and Edgeworth expansions. For bases with an algebraic structure, such as the Hermite polynomial and Fourier bases, the SDM approximation can be chosen so as to satisfy given moment specifications and can be optimized using a quadratic proximity criterion. We apply the SDM approach to the Fokker-Planck-Kolmogorov PDF dynamics of Markov diffusion processes governed by nonlinear stochastic differential equations. This leads to an ordinary differential equation for the SDM dynamics of the approximating PDF. As an example, we consider the Smoluchowski SDE on a multidimensional torus.

## I. INTRODUCTION

Practical solvability of performance analysis and control design problems for stochastic systems often depends on tractability of relevant quantities, such as moments of the system variables. For example, the Kalman filtering and Linear Quadratic Gaussian control theories [1] substantially employ the preservation of Gaussian nature of probability distributions of the state variables governed by linear SDEs. Under the linear dynamics, the first and second order moments of the variables (and more complicated functionals of Gaussian distributions) are amenable to a complete analysis. It is the convenience of linear Gaussian models that makes them so popular in filtering/control with quadratic and related (for example, risk-sensitive) performance criteria. These advantages motivate the approximation of a nonlinear stochastic system by an effective linear model which underlies the stochastic linearization techniques. The latter date back to [3], [5], [12] and have recently been extended to quantum stochastic systems [30].

A different approach to computing the statistical characteristics of a nonlinear stochastic system (oriented at approximating probability distributions rather than system dynamics) consists, for example, in representing the probability distribution of its state variables as a mixture of Gaussian distributions whose parameters evolve in time. In fact, mixed Gaussian distributions arise as exact posterior probability distributions in the Lainiotis multimodel filter [15], where the conditional Gaussian distributions from partial Kalman filters are weighted by recursively updated posterior probabilities of the corresponding linear models conditioned on the observations. This combination of a bank of Kalman filters with a "mixing" block is a recursive implementation of the Bayesian approach. The important property that the resulting mixture of Gaussian PDFs is a legitimate PDF, which is nonnegative everywhere and satisfies the normalization condition, does not always come with PDF approximations, in general. For example, the truncated Gram-Charlier and Edgeworth expansions [2], based on the Hermite polynomials, are not equipped with this feature, although they provide control over moments or cumulants up to an arbitrary given order.

The aim of the present paper is to outline an approach to the approximation of PDFs by using quadratic forms of weighted complex-valued orthonormal basis functions with positive semi-definite Hermitian matrices of unit trace. These matrices are called stochastic density matrices (SDM) in order to emphasize an analogy (and, at the same time, avoid confusion) with the quantum mechanical density matrices [25]. The SDM approximation leads to a legitimate PDF which is nonnegative everywhere and satisfies the normalization condition. Furthermore, it retains the possibility to control the moments of the PDF for orthonormal bases with an algebraic structure, such as the Hermite polynomial and Fourier bases. The SDM approximation can be optimized by using a proximity criterion for PDFs based on the second-order relative Renyi entropy [23], which leads to a quadratic minimization problem.

This allows the SDM approach to be applied to PDFs of Markov diffusion processes, governed by nonlinear SDEs, by reducing the approximate numerical integration of the Fokker-Planck-Kolmogorov equation (FPKE) [28] to the solution of an ODE for the SDM, which resembles the Galerkin approximations for parabolic PDEs [7]. As an illustration, we consider a Smoluchowski SDE [8], [11], [31] on a multidimensional torus, which provides an example of a nonlinear stochastic system with rotational degrees of freedom. The SDM approach admits a real version in the case of real-valued basis functions, with an appropriate reformulation of the results. It is relevant to mention a connection of this approach with the methods using the sum of squares (SOS) of polynomials for Lyapunov stability analysis and global optimization [16], [22]. However, the SDM approach, proposed in the present paper, serves a different purpose here and is not restricted to polynomials.

[†]UNSW Canberra, Australia. `igor.g.vladimirov@gmail.com`.

The paper is organised as follows. Section II describes the class of PDFs generated by an SDM and a set of orthonormal functions. Section III relates the algebraic structure of the orthonormal basis to the moments of such PDFs. Section IV discusses effective parameters of the SDM which control the PDF. Section V specifies this class of PDFs for the multivariate Hermite polynomial and Fourier bases. Section VI describes the SDM approximation of a given PDF using a quadratic criterion. Section VII extends the SDM approximation to PDF dynamics of Markov processes. Section VIII considers the Smoluchowski SDE on a multidimensional torus. Section IX reformulates the corresponding FPKE in the spatial frequency domain. Section X describes the SDM approximation of the PDF dynamics using the Fourier basis and provides numerical results. Section XI makes concluding remarks.

## II. STOCHASTIC DENSITY MATRIX

Suppose $G$ is a finite-dimensional state space of a dynamical system. To be specific, we assume that $G$ is a domain in $\mathbb{R}^n$, or an $n$-dimensional torus $\mathbb{T}^n$. In what follows, we use a complex Hilbert space

$$\mathscr{H} := \mathscr{L}^2(G, \nu) := \left\{ f : G \to \mathbb{C} : \int_G |f(x)|^2 \nu(x) \mathrm{d}x < +\infty \right\} \tag{1}$$

of square integrable complex-valued functions on the set $G$ with a weight $\nu : G \to (0, +\infty)$. The norm $\|f\|_{\mathscr{H}} := \sqrt{\langle f, f \rangle_{\mathscr{H}}}$ in the space $\mathscr{H}$ is generated by the inner product

$$\langle f, g \rangle_{\mathscr{H}} := \int_G \overline{f(x)} g(x) \nu(x) \mathrm{d}x, \tag{2}$$

where the integral is over the standard $n$-variate Lebesgue measure, and $\overline{(\cdot)}$ is the complex conjugate. Furthermore, we denote by $\langle X, Y \rangle := \mathrm{Tr}(X^* Y)$ the Frobenius inner product [9] of complex matrices $X$ and $Y$, which reduces to $\mathrm{Tr}(XY)$ for complex Hermitian matrices (the real Hilbert space of such matrices of order $r$ is denoted by $\mathbb{H}_r$, with $\mathbb{H}_r^+$ the corresponding set of positive semi-definite matrices). Also, $(\cdot)^* := \overline{((\cdot))}^{\mathrm{T}}$ is the complex conjugate transpose, and vectors are organised as columns unless specified otherwise. Suppose $(\varphi_k)_{k \in \Omega} \in \mathscr{H}$ is a fixed but otherwise arbitrary orthonormal basis in the Hilbert space (1), whose elements are indexed by a denumerable set $\Omega$. The latter is assumed to be a subset of a multidimensional integer lattice which contains the origin: $0 \in \Omega$. Therefore,

$$\langle \varphi_j, \varphi_k \rangle_{\mathscr{H}} = \delta_{jk}, \qquad j, k \in \Omega, \tag{3}$$

where $\delta_{jk}$ is the Kronecker delta. Now, suppose $\Lambda$ is a given nonempty finite subset of $\Omega$ consisting of $N := \#\Lambda$ elements. Also, let $S := (s_{jk})_{j,k \in \Lambda} \in \mathbb{H}_N^+$ be a matrix of unit trace $\mathrm{Tr}\,S := \sum_{k \in \Lambda} s_{kk} = 1$. Such matrices form a convex subset of the space $\mathbb{H}_N$ which we denote by

$$\mathfrak{S}_N := \left\{ S \in \mathbb{H}_N^+ : \mathrm{Tr}\,S = 1 \right\}. \tag{4}$$

With the orthonormal basis $(\varphi_k)_{k \in \Omega}$ in $\mathscr{H}$ and the set $\Lambda$ being fixed, we associate with $S \in \mathfrak{S}_N$ a real-valued function $p : G \to \mathbb{R}$ defined as a quadratic form

$$p(x) := \nu(x) \sum_{j,k \in \Lambda} s_{jk} \overline{\varphi_j(x)} \varphi_k(x) = \nu(x) \Phi(x)^* S \Phi(x)$$
$$= \nu(x) \langle S, \Phi(x) \Phi(x)^* \rangle, \qquad x \in G, \tag{5}$$

where $\Phi : G \to \mathbb{C}^N$ is a vector-valued map formed from the basis functions as

$$\Phi(x) := (\varphi_k(x))_{k \in \Lambda}. \tag{6}$$

*Lemma 1:* For any matrix $S \in \mathfrak{S}_N$ from the set (4), the corresponding function $p$, defined by (5) and (6), is a PDF on the set $G$.

*Proof:* The positive semi-definiteness $S \succcurlyeq 0$ and the positiveness $\nu > 0$ in (5) imply that $p(x) \geqslant 0$ for all $x \in G$. Furthermore, the orthonormality (3) and the assumption $\mathrm{Tr}\,S = 1$ imply that the function $p$ satisfies the normalization condition $\int_G p(x) \mathrm{d}x = \sum_{j,k \in \Lambda} s_{jk} \langle \varphi_j, \varphi_k \rangle_{\mathscr{H}} = \mathrm{Tr}\,S = 1$, and hence, $p$ is indeed a PDF. ∎

Due to its properties and the role in the construction (5) of a legitimate PDF, the matrix $S$ resembles the density matrices in quantum mechanics [21], [25]. In order to reflect this analogy (and, at the same time, avoid confusion) with the quantum mechanical density matrices and the stochastic matrices of transition probabilities of classical Markov chains [26], we will refer to $S$ as a *stochastic density matrix* (SDM). Since the set $\mathfrak{S}_N$ of SDMs $S$ in (4) is convex and the PDF $p$ in (5) depends linearly on $S$, the set of such PDFs is also convex. If the Hilbert space in (1) is restricted to real-valued functions, then the SDM becomes a real symmetric matrix, and the results that follow can be appropriately reformulated for this case.

## III. ALGEBRAIC STRUCTURE AND MOMENTS

For what follows, we assume that $\varphi_0 = 1$, and hence, in view of (3), the weighting function $\nu$ satisfies $\int_G \nu(x)\mathrm{d}x = 1$ and is itself a PDF. This allows the inner product (2) to be represented as $\langle f, g \rangle_{\mathscr{H}} = \mathbf{E}_\nu(\overline{f}g)$, where $\mathbf{E}_\nu(\cdot)$ denotes the expectation over the PDF $\nu$. Also, suppose the basis functions are algebraically closed in the sense that there exist complex constants $e_{jk\ell}$ such that for any $j, k \in \Omega$, the representation

$$\varphi_j(x)\overline{\varphi_k(x)} = \sum_{\ell \in \Omega} e_{jk\ell}\varphi_\ell(x) \tag{7}$$

holds for all $x \in G$ and contains only a finite number of terms with nonzero coefficients $e_{jk\ell}$. The coefficients $e_{jk\ell}$ in (7), which quantify the algebraic structure of the orthonormal basis $(\varphi_k)_{k \in \Omega}$, can be computed as

$$e_{jk\ell} = \langle \varphi_\ell, \varphi_j\overline{\varphi_k} \rangle_{\mathscr{H}} = \mathbf{E}_\nu(\varphi_j\overline{\varphi_k}\varphi_\ell) \tag{8}$$

and are symmetric with respect to the subscripts $k$, $\ell$. Furthermore, $e_{jk\ell}$ is real and invariant under arbitrary permutations of its subscripts in the real case mentioned at the end of Section II. For any $\ell \in \Omega$, we define the $\ell$th *structure matrix* $E_\ell := (e_{jk\ell})_{j,k \in \Lambda} \in \mathbb{C}^{N \times N}$ by

$$E_\ell := (e_{jk\ell})_{j,k \in \Lambda} = \mathbf{E}_\nu(\overline{\varphi_\ell}\Phi\Phi^*), \tag{9}$$

where the expectation applies entrywise, and the map $\Phi$ is given by (6). In particular, since $\varphi_0 = 1$, then (8) implies that $e_{jk0} = \delta_{jk}$ for all $j, k \in \Lambda$, and hence, $E_0 = I_N$ is the identity matrix of order $N$. Since the left-hand side of (7) is the $(j,k)$th entry of the matrix $\Phi(x)\Phi(x)^*$, the algebraic closedness property is representable in a vector-matrix form:

$$\Phi(x)\Phi(x)^* = \sum_{\ell \in \mho} \varphi_\ell(x)E_\ell. \tag{10}$$

This series contains only a finite number of terms since the matrix $E_\ell$ in (9) vanishes for all but finitely many indices $\ell \in \Omega$ which form the set

$$\mho := \{\ell \in \Omega : E_\ell \neq 0\}. \tag{11}$$

This set, consisting of $L := \#\mho$ elements, depends on the set $\Lambda$ and contains 0 (since $E_0 = I_N$). Substitution of (10) into the definition of the PDF $p$ in (5) leads to

$$\begin{aligned} p(x) &= \nu(x)\sum_{\ell \in \mho} \langle S, E_\ell \rangle \varphi_\ell(x) \\ &= \nu(x)\sum_{\ell \in \mho} \langle E_\ell, S \rangle \overline{\varphi_\ell(x)} = \nu(x)C(S)^*\Psi(x). \end{aligned} \tag{12}$$

Here, $C : \mathbb{H}_N \to \mathbb{C}^L$ is a linear vector-valued map defined in terms of the nonzero structure matrices by

$$C(S) := (\langle E_\ell, S \rangle)_{\ell \in \mho}, \tag{13}$$

and, similarly to (6), the map $\Psi : G \to \mathbb{C}^L$ is formed from the basis functions as

$$\Psi(x) := (\varphi_\ell(x))_{\ell \in \mho}. \tag{14}$$

The second equality in (12) follows from $p$ and $\nu$ being real-valued. In view of this equality and (3), the expectations of the basis functions over the PDF $p$ (which play the role of generalized moments of this PDF) are computed as

$$\mathbf{E}_p\varphi_m := \int_G p(x)\varphi_m(x)\mathrm{d}x = \sum_{\ell \in \mho} \langle E_\ell, S \rangle \langle \varphi_\ell, \varphi_m \rangle_{\mathscr{H}} = \langle E_m, S \rangle \tag{15}$$

for any $m \in \Omega$ and vanish for all $m \in \Omega \setminus \mho$ in view of (11). Therefore, (12) describes the series expansion of the PDF $p$ from (5) over the orthonormal basis in $\mathscr{H}$, and the map $C$ in (13) encodes the dependence of all the nonzero moments in (15) on the SDM $S$. The fact that the moments depend linearly on $S$ allows the SDM to be chosen so as to satisfy given moment specifications for the corresponding PDF $p$ in (5) . Note that the moments $\mathbf{E}_p\varphi_m$, with $m \neq 0$, are responsible for the deviation of the PDF $p$ from the weighting function $\nu$. In particular, (12) implies that the second-order relative Renyi entropy [23] of $p$ with respect to $\nu$ (with the latter playing the role of a reference PDF) can be expressed in terms of the moments in (15) as

$$\begin{aligned} \mathbf{R}(p \| \nu) &:= \ln \int_G \frac{p(x)^2}{\nu(x)}\mathrm{d}x = \ln \mathbf{E}_\nu\left(\left(\frac{p}{\nu}\right)^2\right) \\ &= \ln \sum_{\ell,m \in \mho} \langle E_\ell, S \rangle \langle S, E_m \rangle \langle \varphi_\ell, \varphi_m \rangle_{\mathscr{H}} \\ &= \ln\left(1 + \sum_{\ell \in \mho \setminus \{0\}} |\langle E_\ell, S \rangle|^2\right). \end{aligned} \tag{16}$$

Here, the orthonormality condition (3) is used in combination with the property that $\langle E_0, S \rangle = \langle I_N, S \rangle = \mathrm{Tr}\,S = 1$.

# IV. EFFECTIVE PARAMETERS

Note that the PDF $p$, defined by (5), can also be represented in the form

$$p(x) = \sum_{j \in \Lambda} \sigma_j \left| \sqrt{v(x)}\, \theta_j(x) \right|^2 \tag{17}$$

which resembles the SOS structures [16], [22]. Here, $\sigma_j$ denote the eigenvalues of the SDM $S$ (they are all real and nonnegative and satisfy $\sum_{j \in \Lambda} \sigma_j = 1$), and the functions $\theta_j : G \to \mathbb{C}$ are obtained by a unitary transformation of the functions $\varphi_k$ as

$$\theta_j(x) := \sum_{k \in \Lambda} \overline{u_{kj}}\varphi_k(x), \qquad j \in \Lambda, \ x \in G. \tag{18}$$

Also, $U := (u_{jk})_{j,k \in \Lambda} \in \mathbb{C}^{N \times N}$ is a unitary matrix whose columns are the eigenvectors of $S$, and hence, $S = U\Sigma U^*$, where $\Sigma := \text{diag}_{j \in \Lambda}(\sigma_j)$ is a diagonal matrix formed from the eigenvalues of $S$. The functions $\theta_j$, defined by (18), are also orthonormal elements of the Hilbert space $\mathscr{H}$ in (1), that is, $\langle \theta_j, \theta_k \rangle_{\mathscr{H}} = \delta_{jk}$. The representation (17) of the PDF $p$ as a convex combination of the squares of the functions $\sqrt{v}|\theta_j|$ involves not only the freedom of varying the coefficients $\sigma_j$, but also the possibility of "mixing" the $N$ elements $\varphi_k$ of the given orthonormal basis of $\mathscr{H}$ in an arbitrary unitary fashion (18). Since the SDM $S$ is Hermitian and has unit trace, then the total number of real parameters which enter the PDF $p$ through $S$ is $\dim \mathbb{H}_N - 1 = N^2 - 1$. However, not all of these parameters are active, in general. From the decomposition $S = \Pi(S) + \Pi^\perp(S)$ of the matrix $S$ into the orthogonal projections $\Pi(S)$ and $\Pi^\perp(S) := S - \Pi(S)$ onto the linear subspace

$$\mathscr{E} := \text{span}\{E_\ell : \ \ell \in \mho\}, \tag{19}$$

spanned by the nonzero structure matrices in (9), and its orthogonal complement $\mathscr{E}^\perp$, respectively, it follows that $\Pi^\perp(S)$ does not influence the PDF $p$ in (12). Here, $\mho$ is a finite subset of $\Omega$ given by (11). Hence, the number of effective parameters of $p$ does not exceed the dimension $\dim \mathscr{E}$ of the subspace (19), that is, the rank of an appropriate Gram matrix:

$$\dim \mathscr{E} = \text{rank}(\langle E_\ell, E_m \rangle)_{\ell, m \in \mho}. \tag{20}$$

The following lemma provides a condition for the nonsingularity of the Gram matrix, that is, the linear independence of the structure matrices $E_\ell$, with $\ell \in \mho$, in the Hilbert space $\mathbb{C}^{N \times N}$. To this end, for any finite subset $M \subset \Omega$, we denote by

$$\mathscr{H}_M := \text{span}\{\varphi_k : \ k \in M\} \tag{21}$$

the subspace of $\mathscr{H}$ spanned by the corresponding set of basis functions. Also, let $\mathscr{Q}_\Lambda$ be a set of functions on $G$ associated with the subspace $\mathscr{H}_\Lambda$ as

$$\mathscr{Q}_\Lambda := \{f\overline{g} : \ f, g \in \mathscr{H}_\Lambda\}. \tag{22}$$

From the algebraic closedness (7) of the basis functions, it follows that

$$\mathscr{Q}_\Lambda \subset \mathscr{H}_\mho. \tag{23}$$

Indeed, in view of the notation (21), any two functions $f, g \in \mathscr{H}_\Lambda$ are representable as $f = a^*\Phi$ and $g = b^*\Phi$ for some vectors $a, b \in \mathbb{C}^N$, where $\Phi$ is given by (6). Therefore, (10) implies that $f\overline{g} = a^*\Phi\Phi^*b = \sum_{\ell \in \mho} a^* E_\ell b \varphi_\ell \in \mathscr{H}_\mho$, whence the inclusion (23) follows in view of the arbitrariness of the functions $f, g \in \mathscr{H}_\Lambda$.

*Lemma 2:* The linear subspace $\mathscr{E}$, defined by (19), has full dimension $\dim \mathscr{E} = L$ if and only if the orthogonal complement of the set $\mathscr{Q}_\Lambda$, given by (22), to the subspace $\mathscr{H}_\mho$ consists of the zero function:

$$\mathscr{Q}_\Lambda^\perp \bigcap \mathscr{H}_\mho = \{0\}. \tag{24}$$

*Proof:* As mentioned above in regard to (20), the property $\dim \mathscr{E} = L$ is equivalent to the linear independence of the structure matrices $E_\ell$, with $\ell \in \mho$, in the space $\mathbb{C}^{N \times N}$. Now, a linear combination $\sum_{\ell \in \mho} c_\ell E_\ell$ of these matrices with complex coefficients $c_\ell$ vanishes if and only if so does the quantity $a^* \sum_{\ell \in \mho} c_\ell E_\ell b = \sum_{\ell \in \mho} c_\ell \mathbf{E}_v(\overline{\varphi_\ell} a^*\Phi\Phi^*b) = \mathbf{E}_v(f\overline{g}h)$ for all $a, b \in \mathbb{C}^N$. Here, $f := a^*\Phi \in \mathscr{H}_\Lambda$, $g := b^*\Phi \in \mathscr{H}_\Lambda$, and $h := \sum_{\ell \in \mho} \overline{c_\ell}\varphi_\ell \in \mathscr{H}_\mho$ are auxiliary functions. Hence, the matrices $E_\ell$, with $\ell \in \mho$, are linearly dependent if and only if there exists a function $h \in \mathscr{H}_\mho \setminus \{0\}$ such that $\mathbf{E}_v(f\overline{g}h) = 0$ for all $f, g \in \mathscr{H}_\Lambda$, that is, $h \in \mathscr{Q}_\Lambda^\perp$. This establishes the equivalence between the linear independence of these matrices and the condition (24). ∎

## V. HERMITE POLYNOMIAL AND FOURIER BASES

For completeness, we will now specify the PDFs (5) for two classical Hilbert spaces of functions. Assuming that $G := \mathbb{R}^n$, consider the functions

$$\varphi_k(x) := \frac{H_k(x)}{\sqrt{k!}}, \qquad x := (x_j)_{1 \leqslant j \leqslant n} \in \mathbb{R}^n, \tag{25}$$

which are labeled by $n$-indices $k := (k_j)_{1 \leqslant j \leqslant n} \in \mathbb{Z}_+^n =: \Omega$ (with $\mathbb{Z}_+$ the set of nonnegative integers) and are obtained from the $n$-variate Hermite polynomials

$$H_k(x) := (-1)^{|k|} e^{\frac{1}{2}|x|^2} \partial_x^k e^{-\frac{1}{2}|x|^2}. \tag{26}$$

Here, use is made of the standard multiindex notation $k! := k_1! \times \ldots \times k_n!$, $|k| := k_1 + \ldots + k_n$, and $\partial_x^k := \partial_{x_1}^{k_1} \ldots \partial_{x_n}^{k_n}$, so that $H_k = H_{k_1} \otimes \ldots \otimes H_{k_n}$ is the tensor product of the corresponding univariate Hermite polynomials. The functions (25) form an orthonormal basis in the real Hilbert space $\mathscr{H} := \mathscr{L}^2(\mathbb{R}^n, \nu)$ of real-valued square integrable functions, where the weight $\nu$ is the $n$-variate standard normal PDF

$$\nu(x) := \frac{e^{-\frac{1}{2}|x|^2}}{(2\pi)^{n/2}}, \qquad x \in \mathbb{R}^n. \tag{27}$$

The orthonormal basis $(\varphi_k)_{k \in \mathbb{Z}_+^n}$, given by (25) and (26), satisfies a real-valued version of (7). In view of [10, Eq. (3.13) on p. 28] and [19, Theorem 3.2.1 on p. 13], the structure coefficients (8) are computed as

$$e_{jk\ell} = \frac{\sqrt{j!k!\ell!}}{(m-j)!(m-k)!(m-\ell)!}, \qquad m := \frac{j+k+\ell}{2}, \tag{28}$$

for all $n$-indices $j, k, \ell \in \mathbb{Z}_+^n$ whose entries do not exceed the corresponding entries of $m$ and such that $j + k + \ell$ consists of even integers, with $e_{jk\ell} = 0$ otherwise. In this case, the PDF $p$ in (5) is specified by a real symmetric SDM $S$, and its representation (12) in terms of the structure matrices from (9) takes the form

$$p(x) = \frac{e^{-\frac{1}{2}|x|^2}}{(2\pi)^{n/2}} \sum_{j,k \in \Lambda} s_{jk} \frac{H_j(x)H_k(x)}{\sqrt{j!k!}} = \frac{e^{-\frac{1}{2}|x|^2}}{(2\pi)^{n/2}} \sum_{\ell \in \mho} \langle S, E_\ell \rangle \frac{H_\ell(x)}{\sqrt{\ell!}} \tag{29}$$

in view of (27). Here, $\Lambda$ is a finite subset of $\mathbb{Z}_+^n$, and, in view of (28), the corresponding set $\mho$ in (11) is the Minkowski sum of the set $\Lambda$ with itself: $\mho = \Lambda + \Lambda := \{j + k : j, k \in \Lambda\}$. Although (29) is organised as a truncated Hermite polynomial expansion, its coefficients $\langle S, E_\ell \rangle$ are parameterized by the SDM $S$ in a specific way, which, according to Lemma 1, guarantees that $p$ is a legitimate PDF. As another example, consider the Fourier basis on the $n$-dimensional torus $\mathbb{T}^n$ (where $\mathbb{T}$ is realised as a half-open interval $[0, 2\pi)$) consisting of the functions

$$\varphi_k(x) := e^{ik^{\mathrm{T}}x}, \qquad x \in \mathbb{T}^n, \ k \in \mathbb{Z}^n. \tag{30}$$

These functions are indexed using the $n$-dimensional integer lattice $\mathbb{Z}^n$ and form an orthonormal basis in the complex Hilbert space $\mathscr{H} := \mathscr{L}^2(\mathbb{T}^n, (2\pi)^{-n})$ with a constant weight $(2\pi)^{-n}$. The latter is the PDF of the uniform distribution over the torus. Since the functions (30) satisfy $\varphi_j \overline{\varphi_k} = \varphi_{j-k}$, the algebraic structure coefficients in (7) are given by

$$e_{jk\ell} = \delta_{j-k,\ell}, \qquad j, k, \ell \in \mathbb{Z}^n. \tag{31}$$

In this case, in view of (12), the PDF $p$ in (5) takes the form of a trigonometric polynomial

$$p(x) = \frac{1}{(2\pi)^n} \sum_{j,k \in \Lambda} s_{jk} e^{i(k-j)^{\mathrm{T}}x} = \frac{1}{(2\pi)^n} \sum_{\ell \in \mho} \langle S, E_\ell \rangle e^{i\ell^{\mathrm{T}}x}, \tag{32}$$

where the set $\mho = \Lambda - \Lambda := \{j - k : j, k \in \Lambda\}$ (which is the Minkowski difference of the set $\Lambda$ with itself) is symmetric about the origin in $\mathbb{Z}^n$. Again, Lemma 1 ensures that (32) describes a legitimate PDF on the torus for any SDM $S$. In the multivariate case being considered, the trigonometric polynomial $p$ in (32) is not necessarily reducible to the squared absolute value of a single trigonometric polynomial. The Fejer-Riesz theorem [24] guarantees such a spectral factorization of a nonnegative trigonometric polynomial only in the univariate case. From the representation (17), with $\nu = (2\pi)^{-n}$, it follows that (32) is a mixture of $N$ such factorizations:

$$p(x) = (2\pi)^{-n} \sum_{j \in \Lambda} \sigma_j |\theta_j(x)|^2, \tag{33}$$

where, in accordance with (18), $\theta_j(x) := \sum_{k \in \Lambda} \overline{u_{kj}} e^{ik^{\mathrm{T}}x}$ are trigonometric polynomials. The PDF $p$ in (33) reduces to a single Fejer-Riesz spectral factorization in the case of a rank-one SDM $S = uu^*$, where $u \in \mathbb{C}^N$ is a unit complex vector. Recalling the analogy with the quantum mechanical density matrices mentioned above, such SDMs correspond to pure quantum states [25] which are extreme points of the convex set of mixed states. Therefore, this multivariate Fourier basis example shows that the SDM approach is, in principle, able to produce a wider class of PDFs than those obtained by squaring a single linear combination of elementary functions.

# VI. QUADRATICALLY OPTIMAL SDM APPROXIMATION OF PDFS

We will now consider a problem of approximating a given PDF $f : G \to \mathbb{R}_+$ by the PDF $p$ from (5). More precisely, assuming that the orthonormal basis in $\mathscr{H}$ and the set $\Lambda$ are fixed, the SDM $S$ is varied over the set (4) so as to minimize a discrepancy between the actual PDF $f$ and the approximating PDF $p_S := p$ (which is parameterized by $S$):

$$D(f, p_S) \longrightarrow \min, \qquad S \in \mathfrak{S}_N. \tag{34}$$

For example, one of such proximity criteria is described by a quadratic functional

$$\begin{aligned} D(f,p) &:= \frac{1}{2} \left\| \frac{f-p}{v} \right\|_{\mathscr{H}}^2 = \frac{1}{2} \int_G \frac{(f(x)-p(x))^2}{v(x)} \mathrm{d}x \\ &= \frac{1}{2} \left( \mathrm{e}^{\mathbf{R}(f\|v)} + \mathrm{e}^{\mathbf{R}(p\|v)} \right) - \int_G \frac{f(x)p(x)}{v(x)} \mathrm{d}x, \end{aligned} \tag{35}$$

where use is made of the Renyi entropy from (16), provided $\mathbf{R}(f\|v)$ and $\mathbf{R}(p\|v)$ are finite, and the $\frac{1}{2}$ factor is introduced for further convenience. A similar, though different, "mean integrated squared error" criterion is employed in the kernel density estimation; see, for example, [4], [6] and the references therein. In view of the linear dependence of $p_S$ on $S$, the quantity $D(f, p_S)$, given by (35), is a convex (but not necessarily strictly convex) quadratic function of $S$, which makes (34) a convex quadratic minimization problem over the convex set of SDMs $\mathfrak{S}_N$. We will therefore consider a "regularised" version of this problem:

$$D(f, p_S) - \mu \ln \det S \longrightarrow \min, \qquad S \in \mathfrak{S}_N, \tag{36}$$

where $\mu > 0$ is an additional parameter which weights the strictly convex function $-\ln \det S$, with the latter also playing the role of a barrier function for the SDM $S$ to avoid singularity and hence, to remain positive definite. The resulting strictly convex minimization problem (36) has a unique solution which is described below. The following theorem employs a positive semi-definite self-adjoint operator $\mathscr{A}$, acting on the Hilbert space $\mathbb{H}_N$ and completely specified by the structure matrices (9) as

$$\mathscr{A}(X) := \sum_{\ell \in \mho} \langle E_\ell, X \rangle E_\ell, \qquad X \in \mathbb{H}_N. \tag{37}$$

We will also need a linear operator $\mathscr{B}$ which maps a PDF $f : G \to \mathbb{R}_+$ to a complex Hermitian matrix

$$\mathscr{B}(f) := \sum_{\ell \in \mho} \mathbf{E}_f \varphi_\ell E_\ell, \tag{38}$$

with the latter being related to the generalized moments of $f$:

$$\mathbf{E}_f \varphi_\ell = \int_G f \varphi_\ell \mathrm{d}x = \left\langle \frac{f}{v}, \varphi_\ell \right\rangle_{\mathscr{H}}. \tag{39}$$

The finiteness of these moments is guaranteed by the assumption $\mathbf{R}(f\|v) < +\infty$ in view of the orthonormality (3) and the Cauchy-Bunyakovsky-Schwarz inequality.

*Theorem 1:* Suppose $f : G \to \mathbb{R}_+$ is a given PDF with finite Renyi entropy in (16): $\mathbf{R}(f\|v) < +\infty$. Then for any given value of the barrier parameter $\mu > 0$, the optimization problem (36), with the proximity criterion (35), has a unique solution $S \succ 0$ which satisfies

$$\mathscr{A}(S) - \mu S^{-1} = \lambda I_N + \mathscr{B}(f). \tag{40}$$

Here, $\lambda \in \mathbb{R}$ is found from the normalization condition $\mathrm{Tr}\, S = 1$, and the operators $\mathscr{A}$ and $\mathscr{B}$ are defined by (37)–(39).

*Proof:* Application of the method of Lagrange multipliers to the strictly convex minimization problem (36) yields the following condition of optimality:

$$\begin{aligned} \partial_S (D(f, p_S) &- \mu \ln \det S - \lambda (\mathrm{Tr}\, S - 1)) \\ &= \partial_S D(f, p_S) - \lambda I_N - \mu S^{-1} = 0. \end{aligned} \tag{41}$$

Here, $\lambda \in \mathbb{R}$ is the Lagrange multiplier associated with the trace constraint in (4), and

$$\partial_S D(f, p_S) = \int_G (p_S - f) \Phi \Phi^* \mathrm{d}x \tag{42}$$

is the Frechet derivative of $D(f, p_S)$ in (35) as a composite function of the SDM $S$, where use is made of the pointwise Frechet derivative $\partial_S p_S = v \Phi \Phi^*$ of the PDF $p_S$ in (5). By combining (10) with the moments of the PDF $p_S$ in (15), it follows that

$$\int_G p_S \Phi \Phi^* \mathrm{d}x = \sum_{\ell \in \mho} \mathbf{E}_p \varphi_\ell E_\ell = \mathscr{A}(S), \tag{43}$$

$$\int_G f \Phi \Phi^* \mathrm{d}x = \sum_{\ell \in \mho} \mathbf{E}_f \varphi_\ell E_\ell = \mathscr{B}(f). \tag{44}$$

Here, the operator $\mathscr{A}$, defined by (37), is the second-order Frechet derivative $\partial_S^2 D(f, p_S)$, and use is made of (38), (39). Substitution of (42)–(44) into (41) leads to (40). The uniqueness of the pair $(\lambda, S)$, satisfying (40) together with $S \succ 0$ and $\mathrm{Tr}\, S = 1$, is ensured by strict monotonicity of $\mathrm{Tr}\, S$ and $\mathscr{A}(S) - \mu S^{-1}$ with respect to $S$ in the sense of the partial ordering on $\mathbb{H}_N$ induced by positive semi-definiteness. Indeed, (43) implies that the operator $\mathscr{A}$ is nondecreasing since $\mathscr{A}(X) = \int_G v\langle X, \Phi\Phi^*\rangle \Phi\Phi^* dx \succcurlyeq 0$ for any $X \in \mathbb{H}_N^+$, while the map $S \mapsto -S^{-1}$ is strictly increasing [9]. ∎

The barrier parameter $\mu > 0$ in (36) controls sensitivity of the optimal SDM $S$ to the PDF $f$ being approximated. In particular, $\lim_{\mu \to +\infty} S = \frac{1}{N} I_N$. At the other extreme, for small values of $\mu$, the SDM $S$ can become nearly singular and highly sensitive to $f$.

## VII. SDM APPROXIMATION OF PDF DYNAMICS FOR MARKOV DIFFUSION PROCESSES

The PDF $p_S$ in (5), with the SDM $S$ computed according to Theorem 1, can be used as a legitimate approximation for PDFs of random processes. More precisely, suppose $f(t, \cdot) : G \to \mathbb{R}_+$ is the time-varying PDF of a $G$-valued Markov diffusion process $\xi(t)$ with an infinitesimal generator $\mathscr{G}$ (so that $\partial_t \mathbf{E}\varphi(t, \xi(t)) = \mathbf{E}_f(\partial_t \varphi + \mathscr{G}(\varphi))$ for any smooth test function $\varphi : \mathbb{R}_+ \times G \to \mathbb{C}$; see, for example, [11]). For what follows, it is assumed that the Renyi entropy $\mathbf{R}(f\|v)$ in (16) remains finite. This can be studied by using the integro-differential relations

$$\partial_t e^{\mathbf{R}(f\|v)} = \partial_t \mathbf{E}_f \frac{f}{v} = \mathbf{E}_f\left(\frac{\partial_t f}{v} + \mathscr{G}\left(\frac{f}{v}\right)\right) = 2\mathbf{E}_f \mathscr{G}\left(\frac{f}{v}\right) \tag{45}$$

which follow (provided $f$ and $v$ are smooth enough) from the PDF dynamics governed by the FPKE

$$\partial_t f = \mathscr{G}^\dagger(f), \tag{46}$$

where the adjoint $\mathscr{G}^\dagger$ is understood in the sense of the standard $\mathscr{L}^2$-space with the unit (or constant) weight. Then the SDM $S$ of the corresponding approximation $p_S$ of $f$ acquires dependence on time, which can be represented in the form of an ODE. The theorem below employs a self-adjoint operator $F_S$ on the space $\mathbb{H}_N$, which is the Frechet derivative of the function of $S$ on the left-hand side of (40):

$$F_S(X) = \mathscr{A}(X) + \mu S^{-1} X S^{-1}, \qquad X \in \mathbb{H}_N. \tag{47}$$

Due to the assumption that $\mu > 0$, the operator $F_S$ is positive definite and strictly increasing for any given $S \succ 0$. These properties of $F_S$ are inherited by its inverse $F_S^{-1}$. Also, let $K$ be a linear operator which maps the PDF $f$ to the complex Hermitian matrix

$$K(f) := \sum_{\ell \in \mho} \mathbf{E}_f \mathscr{G}(\varphi_\ell) E_\ell, \tag{48}$$

where $\mathscr{G}$ is the generator of the Markov diffusion process $\xi$, and use is made of the structure matrices (9).

*Theorem 2:* Suppose the PDF $f(t, \cdot)$ of the underlying Markov diffusion process has finite Renyi entropy $\mathbf{R}(f\|v)$ in (16) at every moment of time $t \geqslant 0$. Then the SDM $S$, associated with $f$ according to Theorem 1, satisfies the ODE

$$\dot{S} = F_S^{-1}(K(f)) - \frac{\mathrm{Tr}\, F_S^{-1}(K(f))}{\mathrm{Tr}\, F_S^{-1}(I_N)} F_S^{-1}(I_N), \tag{49}$$

where the operators $F_S$ and $K$ are defined by (47) and (48).

*Proof:* By taking the time derivative on both sides of (40), it follows that

$$F_S(\dot{S}) = \dot{\lambda} I_N + \partial_t \mathscr{B}(f) = \dot{\lambda} I_N + K(f), \tag{50}$$

where use is made of (47) and a combination of (38) with (48). The invertibility of the operator $F_S$ allows (50) to be solved for $\dot{S}$ as

$$\dot{S} = \dot{\lambda} F_S^{-1}(I_N) + F_S^{-1}(K(f)). \tag{51}$$

Since $F_S$ is strictly increasing, then $F_S^{-1}(I_N) \succ 0$. Hence, $\mathrm{Tr}\, F_S^{-1}(I_N) > 0$, and $\dot{\lambda}$ in (51) can be uniquely found so as to satisfy the condition $\mathrm{Tr}\, \dot{S} = 0$ (which comes from the preservation of $\mathrm{Tr}\, S = 1$ in time):

$$\dot{\lambda} = -\frac{\mathrm{Tr}\, F_S^{-1}(K(f))}{\mathrm{Tr}\, F_S^{-1}(I_N)}. \tag{52}$$

By substituting (52) back into (51), it follows that the SDM $S$ is governed by (49). ∎

The right-hand side of (49) is linear with respect to the PDF $f$ and depends on $S$ in a rational fashion. The latter follows from the representation

$$\mathrm{vec}(F_S^{-1}(X)) = \left(\sum_{\ell \in \mho} \mathrm{vec}(E_\ell)\mathrm{vec}(E_\ell)^* + \mu \overline{S^{-1}} \otimes S^{-1}\right)^{-1} \mathrm{vec}(X) \tag{53}$$

obtained by applying the vectorization $\text{vec}(\cdot)$ of matrices [18] to (37) and (47) and using the relation $S^{\mathrm{T}} = \overline{S}$ in view of $S$ being Hermitian, with $\otimes$ the Kronecker product. If $f$ is an invariant PDF of the Markov process $\xi$, then (48) implies that $K(f) = 0$, and (49) yields $\dot{S} = 0$ in accordance with the static setting considered in Section VI. Now, for nonlinear stochastic systems, the solution $f$ of the FPKE (46) is usually not available in a closed form. In this case (when approximations of $f$ are particularly important), the right-hand side of (49) can be evaluated by replacing the unknown PDF $f$ with its SDM approximation $p_S$, which leads to

$$\dot{S} = F_S^{-1}(Q(S)) - \frac{\text{Tr}\,F_S^{-1}(Q(S))}{\text{Tr}\,F_S^{-1}(I_N)} F_S^{-1}(I_N). \tag{54}$$

Here, $Q$ is a linear operator on the space $\mathbb{H}_N$, obtained by substituting the PDF $p_S$ from (12) into the operator $K$ in (48):

$$Q(S) := K(p_S) = \sum_{m \in \mho} \mathbf{E}_{p_S} \mathscr{G}(\varphi_m) E_m$$

$$= \sum_{\ell, m \in \mho} \langle \varphi_\ell, \mathscr{G}(\varphi_m) \rangle_{\mathscr{H}} \langle E_\ell, S \rangle E_m. \tag{55}$$

The right-hand side of the ODE (54) is a rational function of the entries of $S$ (thus resembling the Riccati equations), which can be represented in the vectorised form by combining (53) with

$$\text{vec}(Q(S)) = \sum_{\ell, m \in \mho} \langle \varphi_\ell, \mathscr{G}(\varphi_m) \rangle_{\mathscr{H}} \,\text{vec}(E_m) \text{vec}(E_\ell)^* \text{vec}(S), \tag{56}$$

following from (55). The constant matrices in (53) and (56) can be precomputed using the structure matrices (9) and the matrix elements of the generator $\mathscr{G}$ over the basis. By construction, the ODE (54) has a positive definite solution whose trace is preserved in time. Therefore, being a closure of (49), the SDM dynamics produce a legitimate approximate solution $p_S$ of the FPKE (46).

## VIII. SMOLUCHOWSKI SDE ON MULTIDIMENSIONAL TORUS

We will now consider an application of the SDM approach to the approximation of PDFs for random processes governed by nonlinear SDEs. As an illustrative example, we will use a version of the Smoluchowski SDE [8], [11], [31] on the $n$-dimensional torus $\mathbb{T}^n$:

$$\mathrm{d}\xi(t) = -\nabla V(\xi(t))\mathrm{d}t + \sigma \mathrm{d}W(t). \tag{57}$$

Here, $\xi := (\xi(t))_{t \geqslant 0}$ is a $\mathbb{T}^n$-valued Markov diffusion process, and $V : \mathbb{T}^n \to \mathbb{R}$ is a twice continuously differentiable function with the gradient $\nabla V := (\partial_{x_k} V)_{1 \leqslant k \leqslant n} : \mathbb{T}^n \to \mathbb{R}^n$, where $\partial_{x_k}(\cdot)$ is the partial derivative with respect to the $k$th angular coordinate. Also, $\sigma > 0$ is a scalar parameter, and $W$ is an $n$-dimensional standard Wiener process. The state space in the form of a multidimensional torus corresponds to systems with rotational degrees of freedom. The SDE (57) (which is understood in the Ito sense) is a noisy version of the gradient descent for the function $V$ and is employed, for example, in the simulated annealing algorithm of stochastic optimization [13]. The function $V$ can be interpreted as the potential energy of a dissipative dynamical system in contact with a heat bath at temperature $T > 0$ which specifies the noise level $\sigma$ in (57) as $\sigma = \sqrt{2k_{\mathrm{B}}T}$, with $k_{\mathrm{B}}$ the Boltzmann constant. This interpretation is motivated by the fact [11] that the invariant probability measure for the Smoluchowski SDE (57) is absolutely continuous with the Gibbs-Boltzmann PDF

$$f_*(x) = \frac{\mathrm{e}^{-\beta V(x)}}{Z(\beta)}, \qquad x \in \mathbb{T}^n, \; \beta := \frac{2}{\sigma^2}. \tag{58}$$

Here, $Z(\beta) := \int_{\mathbb{T}^n} \mathrm{e}^{-\beta V(x)}\mathrm{d}x$ is the statistical mechanical partition function [20] of the auxiliary parameter $\beta = \frac{1}{k_{\mathrm{B}}T}$, associated with the potential $V$. The invariant PDF $f_*$ in (58) is a unique equilibrium point of the FPKE (46) for the PDF $f(t, \cdot) : \mathbb{T}^n \to \mathbb{R}_+$ of the random vector $\xi(t)$, which takes the form

$$\partial_t f = \text{div}(f \nabla V) + \frac{\sigma^2}{2} \Delta f. \tag{59}$$

Here, $\text{div}(\cdot)$ and $\Delta(\cdot)$ are the divergence and Laplace operators over the spatial variables. The corresponding generator $\mathscr{G}$ acts on a smooth test function $\varphi : \mathbb{T}^n \to \mathbb{C}$ as

$$\mathscr{G}(\varphi) = -\nabla V^{\mathrm{T}} \nabla \varphi + \frac{\sigma^2}{2} \Delta \varphi. \tag{60}$$

Although the invariant PDF $f_*$ admits a closed-form representation (58), the time evolution (59) of $f$ towards the equilibrium can be complicated, and it is especially so for multiextremum potentials $V$. For example, in the context of molecular dynamics simulation, such $V$ can represent fractal-like energy landscapes (as functions of the dihedral angles in protein macromolecules) which are considered to play an important role in relaxation phenomena associated with protein folding [17]. This motivates investigation of the PDF dynamics for such FPKEs.

## IX. FPKE IN SPATIAL FREQUENCY DOMAIN

We will now reformulate the FPKE (59) in the spatial frequency domain using the fact that all functions on the torus $\mathbb{T}^n$ are $2\pi$-periodic with respect to the angular variables and (under the assumption of square integrability underlying the Hilbert space $\mathscr{H} := \mathscr{L}^2(\mathbb{T}^n, (2\pi)^{-n})$) can be represented by Fourier series. Let

$$V(x) = \sum_{k \in \mathbb{Z}^n} V_k \varphi_k(x), \qquad f(t,x) = \sum_{k \in \mathbb{Z}^n} f_k(t) \varphi_k(x) \tag{61}$$

be the Fourier series for the potential $V$ and the PDF $f(t, \cdot)$ over the corresponding basis (30). Since the functions $V$ and $f$ are real-valued, their Fourier coefficients satisfy the Hermitian property $V_{-k} = \overline{V_k}$ and $f_{-k} = \overline{f_k}$ for all $k \in \mathbb{Z}^n$ and any time $t$. The normalization condition for the PDF $f(t, \cdot)$ is equivalent to

$$f_0(t) = (2\pi)^{-n} \int_{\mathbb{T}^n} f(t,x) \mathrm{d}x = (2\pi)^{-n}, \qquad t \geqslant 0. \tag{62}$$

Since the potential $V$ is assumed to be twice continuously differentiable, $\|\Delta V\|_\infty := \max_{x \in \mathbb{T}^n} |\Delta V(x)| < +\infty$. Furthermore, if $V$ is infinitely differentiable, it can be shown that the FPKE (59) has a smooth fundamental solution [28]. Similarly to the energy estimates for parabolic PDEs [7], the norm $\|f(t, \cdot)\|_\mathscr{H}$ at any time $t > 0$ is amenable to an upper bound in terms of $\|f(0, \cdot)\|_\mathscr{H}$. This is obtained by applying the Gronwall-Bellman lemma to the following differential inequality which is given here for completeness.

*Lemma 3:* The PDF $f$, governed by the FPKE (59) for the Smoluchowski SDE (57) with an infinitely differentiable potential $V$, satisfies

$$\left(\|f\|_\mathscr{H}^2\right)^{\cdot} \leqslant \left(\|\Delta V\|_\infty - \sigma^2\right)\|f\|_\mathscr{H}^2 + (2\pi)^{-2n}\sigma^2. \tag{63}$$

*Proof:* A combination of (59), (60) with the identities $f \nabla V^{\mathrm{T}} \nabla f = \frac{1}{2}\left(\mathrm{div}(f^2 \nabla V) - f^2 \Delta V\right)$ and $f \Delta f = \frac{1}{2}\Delta(f^2) - |\nabla f|^2$ (and the property that $f$ and $V$ are real-valued) leads to

$$\begin{aligned}
\left(\|f\|_\mathscr{H}^2\right)^{\cdot} &= 2\langle f, \mathscr{G}^\dagger(f)\rangle_\mathscr{H} = 2\langle \mathscr{G}(f), f\rangle_\mathscr{H} \\
&= \langle \Delta V, f^2\rangle_\mathscr{H} - \sigma^2 \|\nabla f\|_\mathscr{H}^2 \\
&\leqslant \|\Delta V\|_\infty \|f\|_\mathscr{H}^2 - \sigma^2 \|g\|_\mathscr{H}^2.
\end{aligned} \tag{64}$$

Here, the first two equalities are, in fact, a particular case of (45) since the Hilbert space $\mathscr{H}$ being considered employs a constant weight. Also, use is made of the Poincare-Wirtinger inequality

$$\begin{aligned}
\|\nabla f\|_\mathscr{H}^2 &= (2\pi)^{-n} \int_{\mathbb{T}^n} |\nabla f|^2 \mathrm{d}x = \sum_{k \in \mathbb{Z}^n \setminus \{0\}} |f_k|^2 |k|^2 \\
&\geqslant \|g\|_\mathscr{H}^2 = \|f\|_\mathscr{H}^2 - f_0^2,
\end{aligned} \tag{65}$$

applied to the function $g := f - f_0$ which satisfies $\int_{\mathbb{T}^n} g \mathrm{d}x = 0$ in view of (62). Substitution of the last equality from (65) into (64) leads to (63). ∎

Another upper bound for the Renyi entropy $\mathbf{R}(f\|v)$ is provided by the inequalities

$$\frac{e^{\mathbf{R}(f\|v)}}{(2\pi)^n \max_{x \in \mathbb{T}^n} f_*(x)} \leqslant e^{\mathbf{R}(f\|f_*)} \leqslant e^{\mathbf{R}(f(0,\cdot)\|f_*)}$$

$$\leqslant \frac{e^{\mathbf{R}(f(0,\cdot)\|v)}}{(2\pi)^n \min_{x \in \mathbb{T}^n} f_*(x)},$$

where both denominators are finite and strictly positive due to (58) and the continuity of $V$ over the torus. The second of these inequalities follows from the dissipation relation

$$\begin{aligned}
\partial_t e^{\mathbf{R}(f\|f_*)} &= 2\mathbf{E}_f \mathscr{G}(h) = 2\mathbf{E}_{f_*}(h\mathscr{G}(h)) \\
&= \int_{\mathbb{T}^n} \mathscr{G}^\dagger(f_*) h^2 \mathrm{d}x - \sigma^2 \mathbf{E}_{f_*}(|\nabla h|^2) \\
&= -\sigma^2 \mathbf{E}_{f_*}(|\nabla h|^2) \leqslant 0,
\end{aligned} \tag{66}$$

whereby $\mathbf{R}(f(t, \cdot)\|f_*)$ is a nonincreasing function of time $t \geqslant 0$. Here, $h := \frac{f}{f_*}$ is the PDF of the Markov process $\xi$ with respect to the invariant measure (so that the first equality in (66) is similar to (45)), and use is made of a change of measure together with the relations $\mathscr{G}^\dagger(f_*) = 0$ and $2h\mathscr{G}(h) = \mathscr{G}(h^2) - \sigma^2 |\nabla h|^2$, with the latter following from (60). Now, (61) allows the matrix elements of the generator $\mathscr{G}$ over the Fourier basis (30) to be computed as

$$\langle \varphi_\ell, \mathscr{G}(\varphi_m)\rangle_\mathscr{H} = V_{\ell-m}(\ell-m)^{\mathrm{T}}m - \frac{\sigma^2}{2}\delta_{\ell m}|m|^2 \tag{67}$$

9

for all $\ell,m \in \mathbb{Z}^n$. Hence, in addition to (62), the Fourier coefficients of the PDF $f$ satisfy a denumerable set of ODEs

$$\dot{f}_j = \left\langle \varphi_j, \mathcal{G}^\dagger(f) \right\rangle_{\mathcal{H}} = \sum_{k \in \mathbb{Z}^n} \left\langle \mathcal{G}(\varphi_j), \varphi_k \right\rangle_{\mathcal{H}} f_k$$

$$= -j^{\mathrm{T}} \sum_{k \in \mathbb{Z}^n} V_k k f_{j-k} - \frac{\sigma^2}{2} |j|^2 f_j, \qquad j \in \mathbb{Z}^n \setminus \{0\}, \tag{68}$$

which represent the FPKE (59) in the spatial frequency domain. Here, the convolution sum comes from the term $f \nabla V$ and is responsible for the coupling of these ODEs (the trivial case of a constant potential $V$ is not considered). The truncation of the set of ODEs (68) (for example, in accordance with Galerkin's method for parabolic PDEs [7]) leads to an approximate system of equations whose solution is not necessarily nonnegative and does not correspond to a legitimate PDF on the torus. This issue can be overcome by using the SDM approximation of PDF dynamics described in Section VII.

## X. SDM APPROXIMATION FOR SMOLUCHOWSKI SDE

In application to the Smoluchowski SDE (57) being considered, the SDM dynamics (54) can be implemented in the vectorised form

$$\mathrm{vec}(S)^{\displaystyle\cdot} = \mathrm{vec}(F_S^{-1}(Q(S))) - \frac{\mathrm{Tr}\, F_S^{-1}(Q(S))}{\mathrm{Tr}\, F_S^{-1}(I_N)} \mathrm{vec}(F_S^{-1}(I_N)) \tag{69}$$

using (53), (56) together with the structure coefficients (31) and the matrix elements (67) over the Fourier basis (30). The corresponding SDM approximation (32) of the PDF $f$ in (59) is

$$p_S(x) = (2\pi)^{-n} C(S)^* \Psi(x), \tag{70}$$

where the maps $C$ in (13) and $\Psi : \mathbb{T}^n \to \mathbb{C}^L$ in (14) are given by

$$C(S) = \left( \sum_{j,k \in \Lambda:\, j-k=\ell} s_{jk} \right)_{\ell \in \mho}, \qquad \Psi(x) = \left( e^{i\ell^{\mathrm{T}}x} \right)_{\ell \in \mho}. \tag{71}$$

Here, $N = (2r+1)^n$ and $L = (4r+1)^n$ in the case when the set $\Lambda$ and the associated set $\mho$ in (11) are the discrete cubes

$$\Lambda := ([-r,r] \textstyle\bigcap \mathbb{Z})^n, \qquad \mho := ([-2r,2r] \textstyle\bigcap \mathbb{Z})^n, \tag{72}$$

with $r$ a positive integer. We will now provide results of a numerical experiment on the SDM approximation of PDFs for the Smoluchowski SDE (57) with $\sigma = 1$ in the two-dimensional case $n = 2$. The potential $V$ was generated as a trigonometric polynomial $V(x) := \sum_{k \in \mathbb{Z}^2:\, |k| \leqslant R} V_k e^{ik^{\mathrm{T}}x} = \sum_{k \in \mathbb{Z}^2:\, |k| \leqslant R} |V_k| \cos(k^{\mathrm{T}}x + \phi_k)$ of $x \in \mathbb{T}^2$ with $R = 5$ (that is, 41 independent harmonics) and exponentially distributed random amplitudes $|V_k| = |V_{-k}|$ and initial phases $\phi_k := \arg V_k = -\phi_{-k}$, with the latter being uniformly distributed over the interval $[0, 2\pi)$ for any $k \neq 0$; see Fig. 1. The SDM approximation (70)
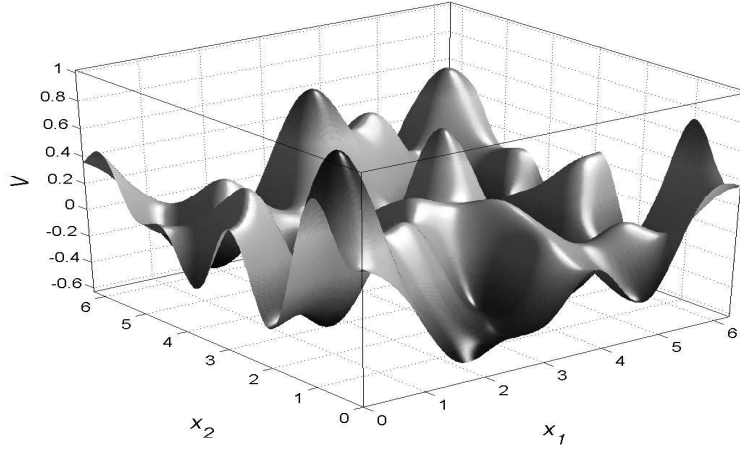


Fig. 1. A multiextremum potential $V$ generated as a trigonometric polynomial of the spatial variables $0 \leqslant x_1, x_2 < 2\pi$ with random coefficients.

was produced using (69) and (71) with the sets (72) with $r = 2$, $N = (2r+1)^2 = 25$, $L = (4r+1)^2 = 81$ and $\mu = 0.01$. The initial conditions for the actual PDF $f$ and the SDM $S$ were $f(0,\cdot) = \frac{1}{4\pi^2}$ and $S(0) = \frac{1}{N}I_N$, with the latter corresponding to the same uniform distribution over the torus with the PDF $p = \frac{1}{4\pi^2}$. The actual PDF $f$ was obtained through the standard finite-difference numerical solution of the FPKE on a mesh of $10^4$ points with time step 0.002. The evolution of the relative values $\frac{D(f,p_S)}{D(f,0)}$ of the quadratic proximity criterion (35) (with $D(f,0)$ being the squared $\mathscr{L}^2$-norm of $f$) is presented in Fig. 2
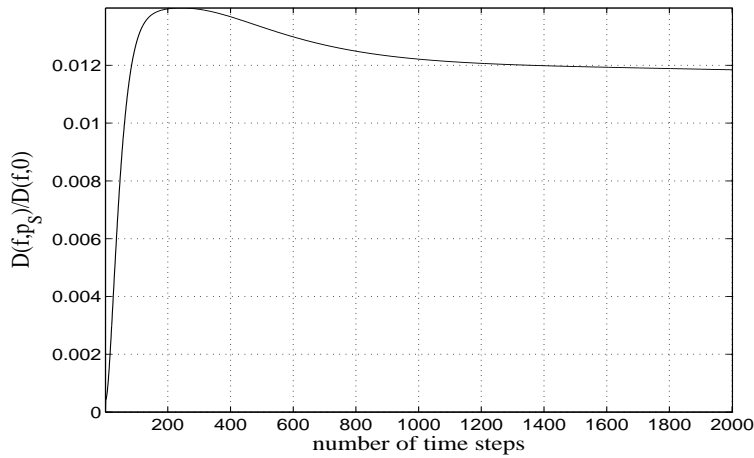
Fig. 2. The relative error $\frac{D(f,p_S)}{D(f,0)}$ of the SDM approximation in the course of evolution of the actual PDF $f$.

which shows that the relative error remained within 1.4%. That the time interval captures the essential part of the movement of the system towards equilibrium is seen from Fig. 3, whereby both the actual PDF $f$ and its SDM approximation $p_S$ at the end of simulation were close to the invariant PDF $f_*$.
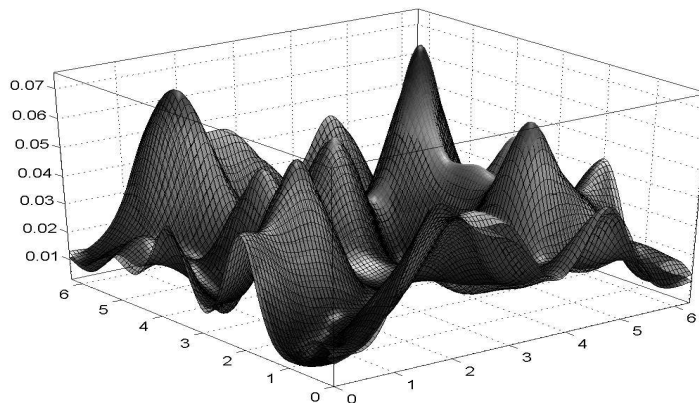


Fig. 3. The invariant PDF $f_*$ (semitranslucent surface), the actual PDF $f$ (black wireframe) and its SDM approximation $p_S$ (red wireframe) after 2000 steps of evolution.

## XI. CONCLUSION

We have outlined an approach to the approximation of PDFs by quadratic forms of weighted orthonormal basis functions parameterised by SDMs. The SDM approximation produces a legitimate PDF which satisfies the normalization condition and is nonnegative everywhere. For orthonormal bases with an algebraic structure, we have provided an optimization procedure for the SDM approximation using a quadratic criterion. The optimal SDM approximation has been extended to PDFs of Markov diffusion processes, and an ODE has been obtained for the SDM dynamics which yields a legitimate approximate solution of the FPKE. This has been demonstrated for the Smoluchowski SDE with a multiextremum potential on a torus. The SDM approach is also applicable (in the spirit of projective filtering [29]) to the approximation of posterior PDFs governed by the Kushner-Stratonovich equations [14]. In this application, the conditional SDM would play a part similar to that of the covariance matrix computed in the Kalman filters. However, unlike the covariance matrices, the SDM is not restricted to the second moments and involves higher order moments of the system variables. Numerical integration of the SDM dynamics can be implemented in a square-root form using the smoothness of the Cholesky factorization of positive definite matrices [27]. Error analysis of the SDM approximation is another line of research to be tackled in future publications.

## REFERENCES

[1] B.D.O.Anderson, and J.B.Moore, *Optimal Control: Linear Quadratic Methods*, Prentice Hall, London, 1989.
[2] O.E.Barndorff-Nielsen, and D.R.Cox, *Asymptotic Techniques for Use in Statistics*, Chapman & Hall, London, 1989.

[3] R.C.Booton, The analysis of nonlinear control systems with random inputs, Proc. Symp. Nonlinear Circuit Analysis, Polytechnic Inst. Brooklyn, New York, vol. 2, 1953.

[4] Z.I.Botev, J.F.Grotowski, and D.P.Kroese, Kernel density estimation via diffusion, *Annals Statist.*, vol. 38, no. 5, 2010, pp. 2916–2957.

[5] T.K.Caughey, Equivalent linearization technique, *J. Acoust. Soc. Amer.*, vol. 35, no. 11, 1963, pp. 1706–1711.

[6] T.Duong, Ks: kernel density estimation and kernel discriminant analysis for multivariate data in R, *J. Statist. Soft.*, vol. 21, no. 7, 2007, pp. 1–16.

[7] L.C.Evans, *Partial Differential Equations*, American Mathematical Society, Providence, 1998.

[8] I.I.Gikhman, and A.V.Skorokhod, *The Theory of Stochastic Processes*, Springer, Berlin, 2004.

[9] R.A.Horn, and C.R.Johnson, *Matrix Analysis*, Cambridge University Press, New York, 2007.

[10] S.Janson, *Gaussian Hilbert Spaces*, Cambridge University Press, Cambridge, 1997.

[11] I.Karatzas, and S.E.Shreve, *Brownian Motion and Stochastic Calculus*, 2nd Ed., Springer-Verlag, New York, 1991.

[12] I.E.Kazakov, Approximate probabilistic analysis of the accuracy of performance of essentially nonlinear control systems, *Automat. Remote Contr.*, vol. 17, 1956, pp. 423–450.

[13] S.Kirkpatrick, C.D.Gelatt, and M.P.Vecchi, Optimization by Simulated Annealing, *Science*, vol. 220, no. 4598, 1983, pp. 671–680.

[14] H.J.Kushner, On the differential equations satisfied by conditional probability densities of Markov processes, with applications, *J. SIAM Control, Ser. A*, vol. 2, no. 1, 1964, pp. 106–119.

[15] D.G.Lainiotis, Partitioned linear estimation algorithms: discrete case, *IEEE Trans. Automat. Contr.*, vol. 20, no. 2, 1975, pp. 255–257.

[16] J.B.Lasserre, Global optimization with polynomials and the problem of moments, *SIAM J. Optim.*, vol. 11, no. 3, 2001, pp. 796–817.

[17] G.Lois, J.Blawzdziewicz, and C.S.O'Hern, Protein folding on rugged energy landscapes: conformational diffusion on fractal networks, *Phys. Rev. E*, vol. 81, 2010, 051907.

[18] J.R.Magnus, *Linear Structures*, Oxford University Press, New York, 1988.

[19] P.Malliavin, *Stochastic Analysis*, Springer, Berlin, 1997.

[20] N.F.G.Martin, and J.W.England, *Mathematical Theory of Entropy*, Addison-Wesley, Reading, Mass., 1981.

[21] P.-A.Meyer, *Quantum Probability for Probabilists*, Springer, Berlin.

[22] P.A.Parrilo, Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization, PhD thesis, California Institute of Technology, Pasadena, California, 2000.

[23] A.Renyi, On measures of entropy and information, Proc. 4th Berkeley Sympos. Math. Statist. Prob., vol. I, 1961, 547–561.

[24] F.Riesz, and B.Sz.-Nagy, *Functional Analysis*, F.Ungar Publishing, New York, 1955.

[25] J.J.Sakurai, *Modern Quantum Mechanics*, Addison-Wesley, Reading, Mass., 1994.

[26] A.N.Shiryaev, *Probability*, 2nd Ed., Springer, New York, 1996.

[27] S.P.Smith, Differentiation of the Cholesky algorithm, *J. Comput. Graph. Statist.*, vol. 4, no. 2, 1995, pp. 134–147.

[28] D.W.Stroock, *Partial Differential Equations for Probabilists*, Cambridge University Press, Cambridge, 2008.

[29] R.Van Handel, and H.Mabuchi, Quantum projection filter for a highly nonlinear model in cavity QED, *J. Opt. B: Quantum Semiclass. Opt.*, vol. 7, 2005, pp. S226–S236.

[30] I.G.Vladimirov, and I.R.Petersen, Gaussian stochastic linearization for open quantum systems using quadratic approximation of Hamiltonians, Proc. MTNS 2012, Melbourne, Victoria, 9–13 July 2012 (preprint: arXiv:1202.0946v1 [quant-ph], 5 February 2012).

[31] R.Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press, New York, 2001.